

CLAIMS

What may be claimed is:

1. A browsable database system for use with biological information, comprising;

at least one datastore of biological sequence data, including at least one of gene sequence data and protein sequence data;

an ontology of categories of biological functions mapped to statistical models trained on families of biological sequences related to the biological functions;

an input receptive of at least one user selection indicating a biological function of said ontology;

a recognizer adapted to identify multiple alignments of biological sequence data based on said sequence datastore and a statistical model related to a function indicated by the user selection; and

an output adapted to communicate the multiple alignments to a user providing the user selection.

2. The system of claim 1, further comprising at least one datastore of curated phylogenetic trees organized into families of sequences based on global sequence similarity, wherein the families are divided into subfamilies according to sequence function and the families and subfamilies are mapped to appropriate statistical models.

3. The system of claim 2, further comprising an output communicating contents of the phylogenetic trees to the user in accordance with user navigation selections.

4. The system of claim 2, further comprising a text searcher receptive of user defined text and adapted to select families and subfamilies of the phylogenetic trees by matching the text to contents of the phylogenetic trees.

5. The system of claim 2, further comprising an input receptive of a user-defined sequence, wherein said recognizer is adapted to select families and subfamilies related to statistical models achieving high scores respective of the user-defined sequence.

6. The system of claim 1, further comprising an output communicating contents of the ontology to the user in accordance with user navigation selections.

7. The system of claim 1, further comprising a text searcher receptive of user defined text and adapted to select functional categories and subcategories by matching the text to contents of the ontology.

8. The system of claim 1, further comprising an input receptive of a user-defined sequence, wherein said recognizer is adapted to select functional categories and subcategories related to statistical models achieving high scores respective of the user-defined sequence.

9. The system of claim 1, further comprising an input receptive of database selections, wherein said recognizer is adapted to identify sequences in a subset of multiple sequence datastores based on the database selections.

10. The system of claim 1, further comprising an input receptive of a user selection of a Boolean operator, wherein said recognizer is adapted to identify the multiple alignments in accordance with the Boolean operator.

11. A method of operation for use with a browsable biological database, comprising;

communicating an ontology of categories of biological functions to a user, wherein the biological functions are mapped to statistical models trained on families of biological sequences related to the biological functions;

receiving at least one user selection indicating a biological function of the ontology;

accessing at least one sequence datastore of biological sequence data, including at least one of gene sequence data and protein sequence data;

employing pattern recognition to identify multiple alignments of biological sequence data based on contents of the sequence datastore and a statistical model related to a function indicated by the user selection; and

communicating the multiple alignments to the user providing the user selection.

12. The method of claim 11, further comprising communicating at least one set of curated phylogenetic trees to a user, wherein the trees are organized into families of sequences based on global sequence similarity, the families are divided into subfamilies according to sequence function, and the families and subfamilies are mapped to appropriate statistical models.

13. The method of claim 12, further comprising navigating and selecting contents of the phylogenetic trees in accordance with user navigation selections.

14. The method of claim 12, further comprising:
receiving user defined text; and
selecting families and subfamilies of the phylogenetic trees by matching the text to contents of the phylogenetic trees.

15. The method of claim 12, further comprising:
receiving a user-defined sequence; and
selecting families and subfamilies related to statistical models achieving high scores respective of the user-defined sequence.

16. The method of claim 11, further comprising communicating contents of the ontology to the user in accordance with user navigation selections.

17. The method of claim 11, further comprising:
receiving user defined text; and
selecting functional categories and subcategories by matching the text to contents of the ontology.

18. The method of claim 11, further comprising:
receiving a user-defined sequence; and
selecting functional categories and subcategories related to
statistical models achieving high scores respective of the user-defined sequence.
19. The method of claim 11, further comprising:
receiving a set of database selections from the user; and
identifying sequences in a subset of multiple sequence datastores
based on the database selections.
20. The method of claim 11, further comprising:
receiving a user selection of a Boolean operator; and
identifying the multiple alignments in accordance with the Boolean
operator.

21. A method for constructing a browsable database for use with biological information, comprising:

clustering biological sequences into families based on global sequence similarity, wherein the biological sequences include at least one of protein sequences and gene sequences;

aligning the families by generating statistical models based on biological sequence clusters associated with the families; and

dividing the families into subfamilies of sequences sharing a common functional attribute, including at least one of molecular function and biological process.

22. The method of claim 21, further comprising extending an original family to include additional members based on the statistical models.

23. The method of claim 21, further comprising producing family trees based on the alignments.

24. The method of claim 21, further comprising selecting curators based on areas of expertise.

25. The method of claim 21, further comprising employing curators to review and annotate family trees in a distance tree context, wherein a curator links a distance tree of a family to sequence-level annotations related to sequences in the family.

26. The method of claim 21, further comprising providing subfamilies with biologically meaningful names.

27. The method of claim 21, further comprising assigning families and subfamilies to appropriate function and process categories of a biological function ontology.

28. The method of claim 21, further comprising scoring the statistical models against biological sequences.

29. The method of claim 21, further comprising relating biological sequences to functions associated with statistical models achieving high scores respective of those biological sequences.

30. The method of claim 21, further comprising training a statistical model on sequences related to a subfamily.